DOCUMENT RESUME

ED 430 571 IR 057 350

AUTHOR	Bouche, Nicole
TITLE	Digitization for Scholarly Use: The Boswell Papers Project at the Beinecke Rare Book and Manuscript Library.
INSTITUTION	Council on Library and Information Resources, Washington, DC.
ISBN	ISBN-1-887334-66-1
PUB DATE	1999-03-00
NOTE	22p.
AVAILABLE FROM	Council on Library and Information Resources, 1755 Massachusetts Ave., NW, Suite 500, Washington, DC 20036; Web site: http://www.clir.org ($15).
PUB TYPE	Reports - Descriptive (141)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Libraries; Access to Information; *Archives; Higher Education; Information Storage; Information Technology; *Library Collections; Library Services; Optical Scanners; Preservation; Printed Materials
IDENTIFIERS	*Digital Scanning; Digital Technology; Digitizing; *Manuscripts; Scholarly Communication; Special Collections (Library); Yale University CT

ABSTRACT

This paper reports on a project that involved the digitization of manuscripts from the Boswell Collection (i.e., personal papers of James Boswell) by the Beinecke Rare Book and Manuscript Library at Yale University (Connecticut). In this case, the digitization process was designed to serve a group of scholars already at work on a publication series, and so distinguishes itself from many others by its focus on the scholarly communication process rather than on giving broad access to collections through the Internet. The paper discusses the many reasons that a special collections library might undertake a digital conversion program, and shares the staff's insights into how digital technology has found its place in the Beinecke Rare Book and Manuscript Library. Specifically, the paper discusses: background on the library and the Boswell collection; project origins; forming a digitization strategy, including choices regarding where the library should place its emphasis; getting the project underway; the scanning process; results and benefits for the library; results and benefits for the scholar-editors; and conclusions. (AEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 430 571

Digitization for Scholarly Use:

The Boswell Papers Project at The Beinecke Rare Book and Manuscript Library

by Nicole Bouche

March 1999

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

B.H. Leney

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Council on Library and Information Resources

Commission on Preservation and Access

Digital Libraries

Economics of Information

Leadership

BEST COPY AVAILABLE

2

IR057350

Digitization for Scholarly Use:

The Boswell Papers Project at The Beinecke Rare Book and Manuscript Library

by Nicole Bouché

March 1999

Council on Library and Information Resources
Washington, D.C.

About the Author

Nicole Bouché is head of the Manuscript Unit at The Beinecke Rare Book and Manuscript Library at Yale University, where she oversees processing, cataloging, and conservation of the library's manuscript and archival collections and serves as a member of the library's digital library implementation working group. From 1987 to 1993, she was assistant head of the Manuscripts Division of The Bancroft Library at the University of California at Berkeley.

Commission on Preservation and Access

The Commission on Preservation and Access, a program of the Council on Library and Information Resources, supports the efforts of libraries and archives to save endangered portions of their paper-based collections and to meet the new preservation challenges of the digital environment. Working with institutions around the world, CLIR disseminates knowledge of best preservation practices and promotes a coordinated approach to preservation activity.

Digital Libraries

The Digital Libraries program of the Council on Library and Information Resources is committed to helping libraries of all types and sizes understand the far-reaching implications of digitization. To that end, CLIR supports projects and publications whose purpose is to build confidence in, and increase understanding of, the digital component that libraries are now adding to their traditional print holdings.

ISBN 1-887334-66-1

Published by:

Council on Library and Information Resources
1755 Massachusetts Avenue, NW, Suite 500
Washington, DC 20036

Web site at http://www.clir.org

Additional copies are available for $15.00 from the above address. Orders must be prepaid, with checks made payable to the Council on Library and Information Resources.

The paper in this publication meets the minimum requirements of the American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials ANSI Z39.48-1984.

Copyright 1999 by the Council on Library and Information Resources. No part of this publication may be reproduced or transcribed in any form without permission of the publisher. Requests for reproduction for noncommercial purposes, including educational advancement, private study, or research, will be granted. Full credit must be given to both the author and the Council on Library and Information Resources.

Contents

Foreword iv

Background 1

Project Origins 3

Forming a Digitization Strategy 4

The Project Gets Underway 6

The Scanning Process 7

Results and Benefits for the Library 9

Results and Benefits for the Scholar-Editors 10

Conclusions 13

Foreword

Rare book and special collections libraries play an important role in humanities scholarship: they provide access to primary resources that are often rare and physically fragile. Librarians face an ongoing challenge to maintain a balance between making these resources available for consultation in the present, and ensuring their physical integrity and continued access to them over time. Fortunately, not all primary resource-based research requires access to an original document. Often enough, a scholar will find a critical printed edition of a manuscript collection to be a happy substitute for the real thing, one even preferred at times to traveling to a distant library. Printed editions also obviate the need to handle fragile materials and decipher hard-to-read contents, as when a manuscript's paper is degraded or its inks have bled.

Special collections libraries have discovered that digital surrogates, like critical editions, have certain advantages in offering enhanced access to rare or unique items. They can be consulted in remote locations and, through image processing, they may overcome the limitations of poor image quality. Special collections libraries have become a dynamic locus of digital conversion projects, in part because their staff have seen the potential of digitization to provide access to rich intellectual resources that are normally very hard to use.

This paper reports on one such project, the digitization of manuscripts from the Boswell Collection by The Beinecke Rare Book and Manuscript Library at Yale University. The paper is one of a series that the Council on Library and Information Resources is publishing in order to explore strategies for integrating digital technology into the management of library print and media collections. In this case, the digitization process was designed to serve a group of scholars already at work on a publication series, and so distinguishes itself from many others by its focus on the scholarly communication process rather than on giving broad access to collections through the Internet. The paper provides a thoughtful discussion of the many reasons that a special collections library might undertake a digital conversion program, and shares the staff's insights into how digital technology has found its place in The Beinecke Rare Book and Manuscript Library.

Abby Smith
Director of Programs

Background

The Beinecke Rare Book and Manuscript Library is Yale University's principal repository for literary papers and for early manuscripts and rare books in the fields of literature, theology, history, and the natural sciences. For close to 50 years, the Beinecke has been publishing James Boswell's manuscripts. Recently, the library has incorporated digital imaging technology into the editorial preparation of the Boswell volumes and explored the role of digital technology in the scholarly communication process.

The library was established in 1963 through the generosity of the Beinecke family as a separately endowed administrative entity within the Yale University Library system. In addition to its general collection of rare books and manuscripts, the Beinecke houses the Yale Collection of American Literature, the Yale Collection of German Literature, the Yale Collection of Western Americana, and the Osborn Collection. The Beinecke collections afford opportunities for interdisciplinary research in such fields as medieval, Renaissance, and eighteenth-century studies, art history, photography, American studies, the history of printing, and modernism in art and literature. The library serves an international research community but is also used heavily by Yale faculty and students for study and teaching. The library awards a number of research fellowships annually to Yale graduate students and outside scholars and sponsors master classes for Yale students to encourage advanced research in the collections. In addition, the library presents many public programs throughout the year, including exhibitions, scholarly conferences, readings and music performances, and other meetings and receptions, most of which are free and open to the public.

Among its many areas of collecting strength, the library's resources for the study of the British eighteenth century are particularly renowned. Chief among these is the Boswell Collection, which contains the personal papers of James Boswell, the eighteenth-century Scottish lawyer, diarist, and associate of Dr. Samuel Johnson. Boswell's now famous *London Journal, 1762-1763*, travel accounts and other works, including *The Life of Samuel Johnson, LL.D.*, first published in 1791, have long served scholars as rich sources for the study of the legal, social, economic, and cultural life of his age. Besides manuscripts of his writings, Boswell's private papers include approximately 4,000 pieces of his correspondence, his legal papers and records, and a large collection of Boswell's printed works. Accompanying these is an extensive family archive spanning six centuries, beginning with the fifteenth-century Boswells of Fife, which richly doc-

uments over 500 years of Scottish social, legal, agricultural, and economic history.

The library's core collection of Boswell papers was purchased by Yale in 1949 from Lieutenant-Colonel Ralph H. Isham. Funds for the purchase were provided by a grant from the Old Dominion Foundation, established by Yale graduate and benefactor, Paul Mellon (Class of 1929), and through the sale of the publication rights to the McGraw-Hill Company.[1] For many years, the papers were held in the Rare Book Room of the Sterling Memorial Library. They were transferred to the Beinecke when it opened in 1963. Since the 1949 Isham purchase, additional papers have been acquired by the library by gift or purchase from the family and from other sources, including a major addition of family estate records acquired as recently as 1993.[2]

In 1950 the Yale Boswell Editorial Project was launched, under the direction of Frederick A. Pottle, a longtime member of Yale's English faculty, with the goal of publishing Boswell's unpublished writings, including correspondence and manuscripts of many of his most prominent books. Over 40 years later, the work continues. In the Yale Editions of the Private Papers of James Boswell (1950-), 27 volumes have been published, 14 in the trade edition and 13 in the research edition. The project staff includes a general editor, a half-time administrative associate, plus several graduate student assistants, all resident at Yale. The editors for the volumes, however, are specialists in the British eighteenth century, drawn from universities and colleges throughout the United States and Great Britain.

Since its inception, the work of the editorial project has been supported by funding from a variety of sources, including generous grants from the National Endowment for the Humanities (1975 through 1997), a modest royalty income (Yale owns the literary rights to Boswell's unpublished works), and numerous gifts from foundations and individual donors. Additional funding and other support are provided by Yale University, which houses the project in the Sterling Memorial Library, pays a portion of the general editor's salary, and covers various other administrative costs.

[1] For a detailed account of the remarkable discovery of the Boswell papers, long thought to be lost, and their subsequent acquisition by Yale, see Frederick A. Pottle, *Pride and Negligence: The History of the Boswell Papers*, New York: McGraw Hill, 1982. For further information about the collection, consult the finding aid for the Boswell Collection (GEN MSS 89), which is available from the Yale Finding Aid Web site (http://webtext.library.yale.edu/finddocs/fadsear.htm). See also Marion S. Pottle, *Catalogue of the Papers of James Boswell at Yale University*, three volumes, Edinburgh: Edinburgh University Press; New Haven: Yale University Press, 1993; Diane J. Ducharme, "The Rest of the Boswells," *Yale University Library Gazette* 62:1-2 (October 1987); and David Buchanan, *The Treasure of Auchinleck: The Story of the Boswell Papers*, New York: McGraw-Hill, 1974.

[2] The library's principal Boswell holdings are currently grouped as follows: Boswell Collection (GEN MSS 89); Boswell Collection–additions (GEN MSS 150); Boswell Collection Supplement (GEN MSS 153).

Project Origins

The decision to digitize a portion of the Boswell Collection arose out of a proposal from the Boswell project's editorial board to scan its reference set of photostat copies of the Boswell manuscripts and make them available to the editorial teams over the Web. For years these less-than-perfect copies have been housed in the Boswell editorial office, where they are used for the purposes of transcription in cases where the copy is sufficiently legible, and to plan and design the scholarly apparatus (footnotes and other commentary) that accompanies the transcribed texts. Any transcription of the text made from these photostats, however, always required word-by-word verification against the originals by the editor of the specific volume. In cases of highly deteriorated originals, transcription could be attempted only from the original manuscript by the editor or an experienced assistant. The editors, typically not resident at Yale, would have to make periodic research visits to New Haven, at project expense, to transcribe and verify texts.[3]

Dramatic reductions in NEH funding for the project in 1995-96, and the threat that the funding might be cut off altogether due to the changing climate for federal funding in the arts and humanities, prompted the Boswell editors to look for ways to speed up their rate of publication. They wanted to try to ensure completion of at least the in-process works before federal funding dried up completely.[4] Hence the rationale for scanning the photostats: Web access to digital versions of the manuscripts could expedite the basic work of transcription and editing by the far-flung team of editors. If more of their work could be completed offsite, they would be able to focus their research visits to Yale on aspects of the work of transcription and text verification that could be done only from the original manuscripts.

When the proposal to digitize the photostats came to the attention of the director of the Beinecke Library, he recognized it as an opportunity for the library to provide essential support to a Yale-based scholarly project, and as possibly an ideal project through which the library could experiment with digitization and Web distribution of manuscripts. Moreover, as the holding repository at Yale for the Boswell papers, the library had a vested interest in ensuring the quality of any digital version of the papers made available on the Web. Obviously, a digital version derived from an original manu-

[3] The less-than-perfect quality of the photostats made further duplication of them using standard photocopying of little use to the editors. Moreover, there had never been a comprehensive microfilm copy made of the Boswell papers, for either reference or preservation. Presumably the creation of the photostats at some time early in the project provided a form of surrogate for the originals that was more attractive to the editors than a microfilm copy, and apart from the editorial team, there appears to have been little demand for copies. Moreover, a comprehensive filming of the papers would have been enormously expensive: much of the material would have required some form of conservation or other stabilization treatment prior to filming, and until recently the entire collection had not been fully processed and listed.

[4] As of FY 1997/98, NEH funding for the Boswell editorial project had ceased and the project is now seeking alternative sources of funding to complete works in progress.

script would be far preferable to one made from the dog-eared photostats, however well they had served the Boswell staff over the years.

Coincidentally, the prospect of scanning a portion of the Boswell Collection came at a most opportune moment in the library's own thinking about the application of digital technology in a large special collections research library. By the mid-1990s, the digital library age (not to say rage) was in full swing: virtually every library (or so it seemed), and especially special collections repositories, were jumping on the bandwagon. As the library management group contemplated the library's entry into the digital arena, it became clear that for the Beinecke Library, which has both extraordinary range and depth of holdings, the options for digitization were virtually endless. Inevitably, curators, archivists, librarians, and library administrators asked the question, "Given the wide range of possibilities, what is the most appropriate application of digital technology for a special collections library such as this one?"

A key factor in the deliberations (perhaps *the* key factor) was how to define a sustainable digital agenda for the library. The group readily concluded that, whatever benefits the library might be expected to derive from digitizing materials in the collections to enhance support of scholarly research and teaching, or even to improve administration of the collections, the cost of providing digital surrogates would add to the already substantial cost of library operations, even if cost savings could be realized in some services as a result of digitizing portions of the collection.[5]

Forming a Digitization Strategy

With these considerations in mind, the library sought to clarify a sustainable strategy for an ongoing commitment to a digital component within its established operations, including the necessary funding base. Given the enormous range of possibilities and the fact that the library could neither afford nor programmatically justify any attempt to scan everything that might be of potential interest to scholars and students, where ought the library to place its emphasis to yield the maximum benefit? And, indeed, how was it to define that benefit? Should the library

1. focus on digitizing special formats that traditionally have been difficult to describe and to service in a research library setting, such as framed and other works of art? The Beinecke holds

[5] For two excellent, recent studies that document the fiscal and administrative challenges of sustaining both collections and reader services over time in major independent research libraries, see Kevin M. Guthrie, *The New-York Historical Society: Lessons from One Nonprofit's Long Struggle for Survival*, San Francisco: Jossey-Bass, 1996 and Jed. I. Bergman. *Managing Change in the Nonprofit Sector: Lessons from the Evolution of Five Independent Research Libraries*, San Francisco: Jossey-Bass, 1996. Both of these studies were funded by The Andrew W. Mellon Foundation.

many such pieces, chiefly as components of a much larger archive.

2. emphasize audiovisual media, which can be found scattered throughout the collections, and which the library receives in ever-increasing numbers as it acquires the archival records of authors, artists, and other cultural leaders of the twentieth century? Digitization of these formats would solve the library's long-standing problems in providing timely and convenient access to these formats and in ensuring the survival of aural and visual data recorded in these more unstable media.
3. focus resources for digital projects on the conversion of material that primarily supports the research and teaching agenda of Yale University or of Yale-based editorial projects for which Beinecke holds key portions or even the core archive?[6] Opportunities for curriculum or project support abound and the potential demand for such services could be endless.
4. embark on interinstitutional projects such as the Advanced Papyrological Information System (APIS), which seeks to improve dramatically access to a vast body of source material that is essential to the work of a scholarly audience far beyond Yale and that is dispersed in libraries and museums throughout the world, including major holdings at the Beinecke?[7] Although papyrus researchers are one of the Beinecke's key constituencies, in absolute numbers they represent a small fraction of the overall scholarly community and of use of the Beinecke collections, even among Yale users.
5. focus on materials for which a digital image matched with powerful browser software offers opportunities for closer, more detailed inspection of the physical artifact than is possible with the naked eye, without having to resort to highly effective but more costly and less accessible specialized equipment, such as microscopes, X-ray technology, and infrared lights. The Beinecke's rich holdings in literary archives and early manuscripts suggest any number of possibilities for using digital surrogates to provide a more flexible tool for scholarly analysis of these texts and of the circumstances of their creation.

These choices are not necessarily mutually exclusive: the library could envision a digital library strategy that encompassed any or all. Nor did the library expect that scanning a portion of the Boswell manuscripts and delivering them over the Web, chiefly to benefit the

[6] Current Yale scholarly editing projects for which the Beinecke Library holds the core archive are the James Boswell Papers Project and the Jonathan Edwards Papers Project. Significant but less comprehensive holdings also support the work of the Benjamin Franklin Papers Project.

[7] Beinecke is a participant in the APIS project, a multi-institutional cataloging, preservation, and digitization project for papyrus, the goal of which is to provide bibliographic and Web-based image access to key papyrus collections nationally and internationally. The project is supported by a grant from the National Endowment for the Humanities.

Boswell editorial project, would yield the definitive answers to any of these questions. But the library did hope that such a project would provide staff with an opportunity to gain practical experience with the methods, costs, benefits, and pitfalls of scanning manuscript holdings, and perhaps lead to insights that would help to guide the development of a longer range strategy for digital applications at the Beinecke.

Consequently, the library decided to create digital surrogates of a portion of the Boswell manuscripts, selected from a group identified by the editorial board as priority items for its immediate publishing agenda (that is, works in progress, well on the way to completion), and would make the digital files available to the editors via the Web. In return, the Boswell editors would provide the library with feedback on the utility of the images to their work of transcription, text verification, and scholarly editing, which all parties hoped would confirm that digitization has positive implications for both the process of editing and the timeframe in which an edited volume could be completed.

The Project Gets Underway

Several months of preparation were spent conducting the preliminary discussions and defining a digitization strategy, selecting the body of materials to be scanned, testing to establish technical specifications for the scanning, and identifying an appropriate vendor to do the scanning onsite. During the first week of July 1997, staff from the Manuscript Unit at Beinecke, assisted by technicians from the firm microMedia, scanned 958 pages of manuscripts in the Boswell Collection pertaining to Boswell's tour of Scotland in the company of Dr. Samuel Johnson. Boswell later published his account of this trip as *The Journal of a Tour to the Hebrides with Samuel Johnson, LL.D.* The material scanned included Boswell's original travel diary kept during the trip, the working manuscript of the book version, and an assortment of associated working notes and manuscript fragments commonly referred to as the Papers Apart.

The editorial rationale for scanning this particular group of material was driven by very practical considerations: *The Journal of a Tour to the Hebrides*, first published in 1785, is one of Boswell's key works. Production of the Yale edition had been underway for several years and was nearing completion. The editor for the volume, Dr. Peter Baker of the University of Virginia, had limited time to spend onsite at Yale for the time-consuming task of final verification of Boswell's text. If he could complete a large portion of his work offsite, the timetable for completing the volume would presumably be shortened.

For the library, the decision to scan the manuscripts relating to the Hebrides tour was equally pragmatic. On the one hand, the target group was a bibliographic whole: the library would obtain a complete surrogate of this portion of the collection, not a fragmentary assortment of "choice" pieces with little research value and in-

frequent user demand. While the library never anticipated much general interest in the Boswell images, it was nevertheless recognized that a complete work in digital form has potential for use in study and teaching beyond the immediate purposes of the Boswell editors. The amount of material to be scanned also was modest (about 500 leaves or 1,000 page images), amounting to less than one linear foot, and yet involved a variety of manuscript types, papers and inks, and hands. The condition of the manuscript leaves ranged from excellent to poor.

Scanning these documents would provide staff with a range of experience and insights into the challenges of creating usable digital surrogates from manuscript sources of the early modern period. This type of source makes up one of Beinecke's great collecting strengths and the library could envision ongoing demand for digital copies of such sources for research, publication, or use in the classroom and public programs. Although much of the material was in less than fine condition, all was in sufficiently good shape to digitize as found: no extensive conservation treatment was required, although a few especially problematic leaves were encased in Mylar prior to scanning. Finally, none of the material was bound: it could all be scanned quite readily using a standard flatbed scanner. Overall, therefore, this would be a challenging, but eminently accomplishable task within the relatively short timeframe available for completing the project. Although never rigidly defined, this timeframe was understood to be "as soon as possible."

The Scanning Process

As the library's goal was to produce a usable product, not a preservation product per se, digital project staff established specifications for scanning by testing a fairly standard range of resolutions, from 300 to 600 dpi, in high contrast, gray scale, and color. On the basis of a comparison of image quality, retrieval time, and cost, they decided to scan the manuscripts as 400 dpi gray-scale TIFF files and to provide them to the editors in JPEG format (76 percent data compression rate). Scanning in color was rejected after initial testing because the higher cost and larger size of color files were judged to be excessive, relative to the likely research benefits of color output. Color scanning would take much longer and would generate 32 to 48 CDs, instead of 16 CDs using gray scale. Subsequent input from the editors has confirmed the correctness of this decision, although from a purely esthetic point of view everyone agrees that color would have been preferable.

The Boswell manuscripts were scanned on a Microtek ScanMaker III flatbed scanner (36-bit single-pass color flatbed, 4,096 shades of gray.) After running tests, and in the interests of being able to complete the scanning in a timely manner, Manuscript Unit digital project staff decided to scan at two basic focus settings, one each for the two main manuscript types. The original diary leaves (about 5 by 7 inches, or 13 by 18 cm) were all silked, highly yellowed and faded,

showing extensive evidence of water damage and lesser amounts of staining, presumably from other sources. The book manuscript of the tour (leaves 8 by 14 inches, or 21 by 36 cms, and smaller) was in much better condition, showing no evidence of water damage or other deterioration, apart from iron gall ink corrosion. These leaves had not been silked, and so presented a much clearer and cleaner artifact to the scanner.

During the scanning process, each image was previewed briefly to ensure proper centering and cropping (just beyond the edge of the piece), but no other adjustments were made at the image level. Ultimately, only about 10 images had to be retaken: one to replace an entirely corrupt file that could not be read, others to correct minor errors in cropping, and one post-scanning decision to rescan a particularly complex long fragment in order to present it in three images (one of the whole plus two half-page shots). During the scanning and the follow-up process of writing the images to CD, no attempt was made to clean up or otherwise improve the appearance of pages, since the goal was to provide the editors with as faithful a representation of the original document as possible.

Scanning took one week (about seven hours a day), plus one day for set-up and several follow-up days spent verifying the images against the originals, ensuring that the digital files were viable, and flagging a few items for retakes or editing. There were two or three full-time equivalent staff working at all times: one or two technicians from microMedia and one Manuscript Unit staff member who was responsible for handling the manuscript leaves and for ensuring that the proper sequencing of images and directory structure for the files were maintained. The file naming and directory structure were based on the call number / box and folder designations from the Boswell Collection finding aid. In all, 16 CDs containing 956 TIFF images were generated: compressed, these were reduced to a single JPEG compact disk. The total cost was about $8,200 or $8.50 per image. This figure includes the base camera cost and vendor labor cost (total $7,400) plus the cost of generating one set of 16 TIFF CDs at $50 per disk ($800). It does *not*, however, include the considerable salary investment of Beinecke staff time to identify an appropriate scanning vendor, set up the project, prepare the materials, oversee the scanning, and conduct bibliographic and other quality control throughout the process.

Although the library had originally intended to deliver the images to the Boswell editors over the Web, the mechanics of doing this, and of possibly associating the images with the Web version of the Boswell Collection finding aid, were deferred pending full implementation of the Beinecke Digital Library, which by that time was under development, in conjunction with a much more ambitious project of scanning approximately 10,000 images from the library's Public Service photonegative file. Instead, the library decided to provide the Boswell editors with a CD of the images in JPEG format, which they could use on their resident machines. Since that time, further assessment and discussion has clarified the library's under-

standing of the utility of maintaining Web access to less used digital files: the JPEG CD, not the Web, now seems to be the more appropriate method to handle such requests in future, when dealing with a highly specialized group of material packaged for a designated user group. No Web browser or other specific viewing software was specified to the Boswell editors: it seemed more appropriate that they experiment with and select a preferred viewer themselves, and digital project staff hoped that they would provide feedback about their experience with multiple viewers. (In fact, this has not occurred.)

Results and Benefits for the Library

This project was the library's first foray into the realm of manuscripts scanning. It provided an excellent introduction to the methods and to the service and management issues associated with the creation, Web delivery, and uses of digital surrogates for manuscript originals. This experience suggested strategies for an overall digital agenda for the library and an approach to funding that the library believes can be sustained over time, largely from internal resources.

The project provided staff with an excellent opportunity to build a base of knowledge and experience in scanning manuscripts and to gain a better understanding of a number of related activities, including the physical and bibliographic preparation of materials; file and directory naming strategies; the pros and cons of scanning in color vs. gray scale; and the components of the scanning and quality control process. In addition, since the summer of 1994, the library had been participating as an early implementer in the development of the Encoded Archival Description (EAD) SGML encoding standard for archival finding aids.[8] A test set of Boswell images was used to experiment with features of the EAD Document Type Definition that enable hypertext links between digital images of original material and the corresponding sections in an archival finding aid. A test version of the finding aid file, linked to image files, was created and was used quite effectively to demonstrate to staff in the library and to the Boswell editors the potential of this form of electronic linkage between digital image files and the corresponding description and contextual information to be found in the finding aid. Because of delays in bringing up the Beinecke Digital Library, however, this component of the project has not been fully realized. Once the digital library is up, the library plans to return to this question and to explore further this form of linked access.

The library also took advantage of the production of the Boswell image files to explore the question of obtaining preservation microfilm from digital files, as an alternative to direct filming of the origi-

[8] For a fuller description of EAD implementation at Yale, including the Beinecke, see "Implementing EAD in the Yale University Library" by Nicole L. Bouché in *American Archivist* 60:4 (Fall 1997), 408-19. For general information on the EAD initiative, see the recently published double issue of the *American Archivist* on the EAD initiative (60:3-4, Summer-Fall 1997).

nals. Initial tests indicated that a somewhat better quality image on film could be obtained if it was derived from the digital file, rather than from direct filming of the originals. An outside vendor is currently producing the full set of microfilm.

For many in the library, the most interesting question that the Boswell digitization project was intended to explore, albeit in a modest and largely unsystematic fashion, was whether the ability to study these manuscripts in digital form would reveal textual and physical details not otherwise readily apparent to the scholarly researcher. Quite apart from the capacity of a researcher to use the browser technology to zoom in and out and to vary contrasts and settings to improve the legibility of texts clearly present on the page, other interesting possibilities for scholarly inquiry have emerged. Manuscript Unit digital project staff found, for example, that they could retrieve a text left by bleed-through from a facing page, now missing, which exists now only as a shadowy passage of mirror writing on the verso of the surviving leaf. By reversing the image to show white writing on a black field and flipping the image onscreen to undo the mirror effect, the text on the missing leaf was readily deciphered.

While recovery of mirror images and other difficult-to-read text is also possible with photographic methods using filters or other methods, the ability of the individual scholar to retrieve the text at will directly onscreen, and to fix it in digital or paper form for future reference, has obvious appeal and wide potential application in collections of early historical and literary manuscripts, found in abundance in a library like the Beinecke. The library undoubtedly will take into consideration the potential for facilitating scholarly research and discovery when assessing the merits of digitizing any particular group of manuscripts and when deciding whether to underwrite a portion of the costs of digital conversion requests initiated by outside parties.

Results and Benefits for the Scholar-Editors

Once the scanning of the Boswell manuscripts was completed and the JPEG copies were generated from the original TIFF images, digital project staff presented the results to the Boswell editors. They were universally enthusiastic about the quality of the images and the potential for using viewing software to inspect more closely textual and physical aspects of the documents. To no one's surprise, they expressed the hope that they would see many other segments of the Boswell Collection converted to digital format. Their systematic use of the digital files in completing the text verification for the Hebrides tour edition more than bore out their initial positive impressions.

On closer examination, image quality for the pages of the book manuscript and of the associated notes (the Papers Apart), which constituted approximately one-third of the material scanned, was judged to be highly successful. Overall, these originals were in relatively good condition and the paper was clean (that is, not highly

stained, foxed, or deteriorated). Except for the few folders of the Papers Apart, which consisted of an assortment of odd-sized paper scraps, the book manuscript was on paper of nearly uniform size (about 8 by 14 inches). All of these leaves were written in a single color of ink that had withstood the test of time, apart from spotty iron gall deterioration, and had neither bled nor, apparently, faded. Also, the handwriting overall was of good size and quite legible. Consequently, the digital images produced from these leaves were excellent, so good in fact that the editor for the volume determined that he did not need to reinspect the originals to verify the text.

From both the library's and the editors' points of view, this discovery was key. Previously, when Boswell editorial staff had had to rely on the photostats for transcription and initial textual verification, scholarly caution (borne out repeatedly by experience) required word-by-word comparison against the originals in all cases. This time, with high-quality digital surrogates to work from, comparison against the original was determined to be unnecessary for about a third of the manuscripts. The editor saved many hours of work and was able to focus all of his summer 1998 research visit to Yale on verification and comparison against the manuscript of the travel diary, which was in much poorer condition. Still, the editor reported that he checked fewer points in the diary against the original than he would have, had he been working in the traditional fashion from the photostats and typed transcripts, rather than from high-resolution digital copies.

For the remaining two-thirds of the manuscripts, comprising Boswell's original travel diary, scanning was judged to be largely successful: limitations in the utility of the digital images were traceable directly to the highly deteriorated condition of the originals and not to the scanning process or to technical choices made in planning the project. The original diary had sustained significant damage from damp and mold, and many of the now disbound leaves were highly stained. All had yellowed considerably, and most suffered from extensive fading of the inks, bleed-through, or iron gall deterioration. In addition, several inks had been used in writing the diary, reflecting stages of composition and revision, in various hands. These inks had faded or bled over time in a manner that was neither consistent nor predictable, thereby complicating the task of the scholar-editor in establishing an authoritative text. Some pages were even more profoundly stained, though with what is uncertain. In some cases, text loss had occurred, particularly along all four edges of the leaves; in most of these cases, legibility of the remaining text was in some way diminished by the physical state of the material.

Moreover, long before the collection made its way to Yale, virtually all of the leaves had been silked to reinforce and stabilize them. This created a screen or hazing effect on the already damaged and stained surfaces. Because of the poor condition of the manuscript leaves and the effect of the silking, faithfully reproduced in the digital images as intended, verification against the original leaves was deemed necessary, even after viewing of the digital versions. In most

cases, however, it revealed nothing more than what had already been determined from viewing the digital file.

Concerning the library's decision not to scan in color, although the editors initially expressed concern about the decision, after looking closely at the gray-scale digital versions and comparing them to the original manuscripts, the editors concluded that even had color scanning been employed, they would have gained little of real value for their work. There would have been no discernible benefit for the third of the material that consisted of the book manuscript and associated notes. In the case of the original diary, in most instances verification against the original manuscript page was still necessary to confirm or refute preliminary conclusions based on the scanned version. Digital project staff were particularly interested to learn from the editor that in many cases, even consultation with the original was inconclusive: the fading of inks, staining, and other damage, was at times so extreme and erratic and the overall condition of the original leaf so poor, that frequently it was virtually impossible to make conclusive statements about writing sequence or content, even after close inspection of the original. While this does not suggest that one ought to forego verification against an original, it does demonstrate that such verification, while necessary, might simply confirm previous findings. Only more complex and invasive methods (some form of chemical analysis of inks, for example) might provide more conclusive evidence.

After working with the full set of gray-scale images, all parties concluded that the largely esthetic benefits to be obtained by color scanning of the Boswell manuscripts did not outweigh the much greater cost of scanning in color (estimated at two to three times the cost of gray scale per image, mostly in labor and supplies), nor the delivery and retrieval problems that the file size of color images would present for Web, or even local system-based access, even when compressed to JPEG format. Nor would it have been feasible to identify, in advance and in any predictable and consistent manner, which pages would have benefited significantly from color scanning.

At present the editors are working quite successfully from a CD set of the images and it is looking less likely that the Beinecke will attempt to make the Boswell files available to them via the Web as a component of the Beinecke Digital Library. Instead, the library probably will maintain the files as a separate, stand-alone set, which it can either deliver to a user in CD format or mount for a special purpose on its Web server, perhaps linked to the digital library but not a permanent feature of it. If the library eventually decides to provide Web access to the Boswell files on a more permanent basis, it probably would be for less specialized users, and the files would be compressed at rates that probably would probably be too lossy for specialized editorial use, at least for the original diary, where the utmost clarity and completeness of the image are most critical.

Conclusions

It would be premature to say that the library has come to any rock-hard conclusions about the kinds of digitization projects for manuscript materials that the Beinecke will undertake in the future, based on this single foray into manuscripts scanning. Technologies associated with digital conversion, storage and retrieval, and delivery to remote users are subject to rapid change, and the library's own sense of opportunities, options, and scholarly and curriculum-based demand for digital surrogates from the library's collections is still evolving. Nevertheless, it is fair to say that some preliminary criteria have emerged through the work on the Boswell digitization project that are helping to shape the library's sense of likely next steps. The library is also influenced by experience with a scanning project that got underway at the same time as the Boswell project and is still ongoing: the scanning of over 10,000 public service photonegatives and other image material from the collections, comprising a cross-section of the library's visual material resources. Equally germane to the library's thinking is the fact that, thus far, the Beinecke has covered the cost of scanning and other digital library initiatives largely out of its operating funds, as it is likely to do in the future, relying only in part on outside grants. Were the library relying wholly or largely on grants or other forms of external funding, some of the assumptions about digital library priorities and strategies for accomplishing the library's objectives might be different. At the very least, the library's ability to cover a good proportion of these costs out of internal funds gives it greater freedom to define projects that directly promote the library's primary mission: to support advanced scholarly inquiry and the educational mission of Yale University.

It is unlikely that the Beinecke will embark on any large-scale scanning of manuscripts or archival collections. Apart from the esthetic advantages of a digital image, especially a color one, over microfilm or a photocopy, the library sees little scholarly benefit to be gained from comprehensive or even partial scanning of the most heavily used archival collections, which tend to be twentieth-century literary archives. In the vast majority of cases, there is little to be revealed by viewing a digital image that is not already readily apparent from a photocopy or microfilm or from routine inspection of the original. Serious scholarly research still requires consultation of the originals, and the more traditional services of supplying microfilm or selective photocopying to individual researchers, with the patron bearing some portion of the cost for the services received, meet the needs of most of the library's primary clientele.[9] A number of important factors—the up-front cost of scanning, the practical realities of managing thousands or millions of digital files that a single major

[9] Beinecke's preservation microfilming agenda is largely dictated by collection use. At the present time, the library is working against a multiyear backlog of unfilmed archival holdings, but if the filming program is sustained, the library will have on hand comprehensive film sets for those collections most likely to be subject to repeated requests for photocopying. When film is available, it, and not the originals, is used to generate the requested paper or film copies for the researcher.

archive could generate, the uncertainties which persist about the longer-range storage and migration of the digital files, and the various delivery issues for digital images made accessible over the Web—argue against the library making so great an investment in a single collection, let alone several collections, given the relatively light use that such digital images are likely to receive over the Web.

Similarly, a library treasures or American Memory–type of project is equally unlikely. The universe of potential treasures in the library's holdings is too large for any such selection to be really informative. For those who want to learn about the library's collections and view images of some of its featured holdings, there is ample opportunity through the library's home page on the Web. An American Memory–type of project, presenting a compilation of manuscript source material in a field of study in which the library is particularly strong, also has limited appeal, given the scholarly research focus of the library's sense of mission. The necessary selectivity of an assemblage of documents chosen for inclusion in a treasures- or topic-driven digital conversion project renders them of limited use to the scholarly researcher, who requires comprehensive access to the much broader universe of pertinent source material in the library's collections.

In contrast, the type of manuscript scanning projects that the Beinecke *is* likely to undertake would be highly focused projects that directly support scholarly research and teaching by serving a clearly defined user community, such as a class- or department-based curriculum, especially (though not exclusively) one that has ties to Yale. The library is also likely to undertake or support digital conversion of manuscript sources where a case can be made that the study and interpretation of the documents could materially benefit from conversion to digital form, allowing flexibility in viewing and assessing the text in a manner that cannot be derived readily from a good quality microfilm, photocopy, or photographic duplicate, or possibly even from close inspection of the original itself. Both the Boswell scanning project and the APIS papyrus project, which are the two major manuscripts scanning projects that the library has undertaken to date, meet these criteria. In the case of the Boswell project, while the library assumed all associated costs in the interests of a Yale-based program, it did so also because it expected to gain valuable practical experience in a new field of endeavor. In all ways, this was successful. However, such an outlay of library funds on behalf of a highly selective audience, producing a digital product with little appeal or utility beyond that group, can hardly be justified as a matter of routine. The scanning costs alone of this relatively small group of material were not inconsiderable, quite apart from the many hours of library staff time devoted over several months to selecting and preparing the files to be filmed, reviewing the scans, and other tasks.

In future, it is much more likely that the library would undertake a project of this kind only if significant matching funds were available from the primary beneficiaries of the project. Apart from such user-driven considerations, the library will undoubtedly use digital

conversion to generate reference copies of individual pieces that are in fragile condition and for which traditional methods of providing a preservation-use photocopy or microfilm are deemed inadequate or are likely to put the materials at risk.

Rather than embark on ambitious manuscripts scanning projects, the library has chosen to focus its digitization efforts on a broad-based, in-house program to scan its holdings of visual materials, including photographs, works of art on paper, paintings, manuscript illuminations, lantern slides, and three-dimensional objects, in the belief that this will yield a greater return for the largest number of researchers and students who use the Beinecke's collections than would any equivalent amount of manuscripts scanning. Visual formats have never been particularly well served by standard cataloging and descriptive practice, compared with text-based documents, in spite of recent efforts to develop and codify effective descriptive standards. At the Beinecke, as at many other libraries where the bias for textual sources is long standing, photographs, works of art and other non-textual sources have generally received less detailed treatment in finding aids and catalog records. Given the size of remaining archival processing backlogs, the library is unlikely to return at this time to processed collections in order to enhance descriptive control of visual materials in them.

Within the context of a digital library, however, key word searches across a wide range of images and the ability to review thumbnail copies of images provide unprecedented access to individual, known images, and to classes of images. The overall benefits in access, preservation of fragile originals, and overall reader services to be gained from comprehensive scanning of the Beinecke's rich and varied holdings of image material far exceed those achievable from large-scale scanning of the library's manuscript and archival collections.

BEST COPY AVAILABLE

COUNCIL ON LIBRARY AND INFORMATION RESOURCES

1755 MASSACHUSETTS AVENUE, N.W., SUITE 500 WASHINGTON, DC 20036-2124
Telephone: 202-939-4750 • Fax: 202-939-4765 • E-mail: info@clir.org • Web: www.clir.org